

# SCIENTIFIC REPORTS

OPEN

## Intron retention-dependent gene regulation in *Cryptococcus neoformans*

Received: 13 June 2016  
Accepted: 03 August 2016  
Published: 31 August 2016

Sara Gonzalez-Hilarion<sup>1,\*</sup>, Damien Paulet<sup>2,\*</sup>, Kyung-Tae Lee<sup>3</sup>, Chung-Chau Hon<sup>4</sup>, Pierre Lechat<sup>5</sup>, Estelle Mogensen<sup>1</sup>, Frédérique Moyrand<sup>1</sup>, Caroline Proux<sup>2</sup>, Rony Barboux<sup>1</sup>, Giovanni Bussotti<sup>5</sup>, Jungwook Hwang<sup>6</sup>, Jean-Yves Coppée<sup>2</sup>, Yong-Sun Bahn<sup>3</sup> & Guilhem Janbon<sup>1</sup>

The biological impact of alternative splicing is poorly understood in fungi, although recent studies have shown that these microorganisms are usually intron-rich. In this study, we re-annotated the genome of *C. neoformans* var. *neoformans* using RNA-Seq data. Comparison with *C. neoformans* var. *grubii* revealed that more than 99% of ORF-introns are in the same exact position in the two varieties whereas UTR-introns are much less evolutionary conserved. We also confirmed that alternative splicing is very common in *C. neoformans*, affecting nearly all expressed genes. We also observed specific regulation of alternative splicing by environmental cues in this yeast. However, alternative splicing does not appear to be an efficient method to diversify the *C. neoformans* proteome. Instead, our data suggest the existence of an intron retention-dependent mechanism of gene expression regulation that is not dependent on NMD. This regulatory process represents an additional layer of gene expression regulation in fungi and provides a mechanism to tune gene expression levels in response to any environmental modification.

Alternative splicing increases diversity in transcriptome and proteome profiles by generating multiple mRNA isoforms from a single gene<sup>1</sup>. Three main types of alternative splicing have been described: exon skipping (ES), alternative splice site selection (A-SS), and intron retention (IR)<sup>2</sup>. In metazoans, in which most genes contain introns and the vast majority are alternatively spliced, it is well documented that alternative splicing can have a great impact on gene function, tissue-specific expression, and disease<sup>3–5</sup>. In fungi, the biological impact of alternative splicing has been poorly studied in part because most research has been done in *Saccharomyces cerevisiae*, an organism with very few introns (<5% of genes contain introns). Nevertheless, in this model yeast the intron containing genes represent 50% of the total transcripts because they are highly expressed genes and recent analyses of next generation sequencing data revealed numerous cases of alternative splicing<sup>6,7</sup>. In addition, in recent years a large number of fungal genomes have been sequenced and introns appear to be more common than previously anticipated. For example, the percentage of intron-containing genes ranges from 2.4% in *Candida glabrata*, 14.5% in *Yarrowia lipolytica*, and 47% in *Schizosaccharomyces pombe* to >99% in *Cryptococcus neoformans*<sup>8–10</sup>. Moreover, recent reports suggest that alternative splicing is not rare in fungi<sup>11,12</sup>. These studies revealed that different types of alternative splicing occur in varying proportions among the kingdoms. In fungi and plants, IR is the most common type of alternative splicing and ES is very rare, while ES is the prevalent type of alternative splicing in animals<sup>11–14</sup>. Some studies also suggest that alternative splicing plays a role in the regulation of virulence in pathogenic fungi<sup>13,15–17</sup>, though few cases of functional alternative splicing have been reported<sup>18</sup>. At this time, the biological functions of alternative splicing in fungi remain largely unknown.

<sup>1</sup>Institut Pasteur, Unité Biologie des ARN des Pathogènes Fongiques, Département de Mycologie, F-75015, Paris, France. <sup>2</sup>Institut Pasteur, Plate-forme transcriptome et Epigénome, Département Génomes et Génétique, F-75015, Paris, France. <sup>3</sup>Department of Biotechnology, College of Life and Biotechnology, Yonsei University, Seoul 03722, Korea. <sup>4</sup>RIKEN Center for Life Science Technologies, Yokohama Institute, Division of Genomic Technology, Yokohama, 230-0045, Japan. <sup>5</sup>Institut Pasteur, HUB Bioinformatique et Biostatistique, C3BI, USR 3756 IP CNRS, F-75015, Paris, France. <sup>6</sup>Graduate School of Biomedical Science and Engineering, Hanyang University, Seoul, Korea. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to G.J. (email: janbon@pasteur.fr)

The encapsulated basidiomycete pathogenic yeast *C. neoformans* is an intron-rich organism. *C. neoformans* exists as two varieties (var. *grubii* and var. *neoformans*) characterized by distinct epidemiological distribution and pathobiological properties<sup>19</sup>. A recent re-annotation of the *C. neoformans* var. *grubii* genome revealed spectacular complexity of its transcriptome<sup>10</sup>. More than 99% of the genes contain introns, and most are multi-intronic with an average of 5 introns per gene. These introns have been shown to be necessary for gene expression<sup>20</sup>. Moreover, preliminary analyses suggested that all types of alternative splicing are present although, as in other fungi, IR appears to be most common<sup>10</sup>.

In this study, we re-annotated the genome of the pathogenic fungus *C. neoformans* var. *neoformans* using RNA-Seq data. This analysis revealed thousands of alternative splicing events affecting a very large number of genes in this yeast. We show that these alternative splicing events mostly result from intron retention and are tightly and specifically regulated by growth conditions. We also provide evidence demonstrating that alternative splicing does not greatly impact proteome diversity. Instead, our data suggest that intron retention regulation provides a mechanism for *C. neoformans* to tune gene expression levels in response to external cues and may aid the pathogen in survival and proliferation in diverse environmental niches.

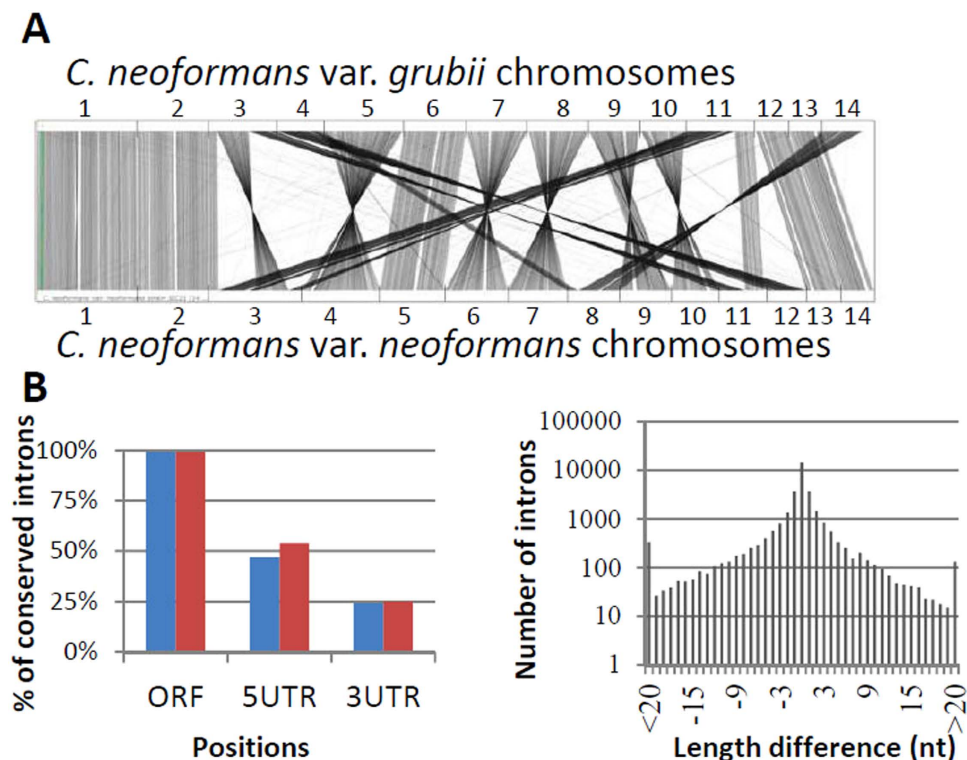
## Results

**Intron positioning in *C. neoformans* var. *neoformans*.** The previous annotation of the *C. neoformans* var. *neoformans* genome largely depended on bioinformatics sequence analysis and comparison with other organisms<sup>21</sup>. Although it is a reasonable strategy to obtain a first draft of a genome annotation, our previous re-annotation of the *C. neoformans* var. *grubii* genome revealed that intron and exon positions are very difficult to predict bioinformatically and that a large proportion of the predicted protein sequences might be wrong when no transcriptomic data is available<sup>10</sup>. Thus, to precisely localize all introns within the *C. neoformans* var. *neoformans* genome, deep-coverage paired-ends strand-specific RNA sequences were generated from 6 different conditions in triplicate. A total of  $1.4 \times 10^9$  strand-specific sequences were aligned to the JEC21 reference genome. Read alignments were compared to the initial gene set of 6,273 predicted coding genes. We found at least 30 reads spanning predicted exon/intron boundaries for 86% of the introns present in the annotation ( $n = 37,814$ ), confirming the *in silico* predicted gene structures. In contrast, 2.3% of the annotated introns had no spanning reads despite being within an expressed gene, suggesting potential incorrect annotations. More notably, we identified 4,782 new introns, resulting in sequence alteration of nearly one-third of the coding sequences ( $n = 2,102$ ). We also identified 497 new coding genes and removed 131 coding genes of the original set, mainly through gene fusion or through re-annotation as pseudogenes ( $n = 31$ ) (Table S1). Overall, 6,639 protein-coding genes and 231 pseudogenes were annotated.

In order to validate our new annotation and to gain some insights concerning the evolutionary conservation of intron positions within the *C. neoformans* varieties, we compared the protein-coding gene structures of the var. *neoformans* with those of the var. *grubii*. First, we used a BLASTp analysis to identify 370 proteins of *C. neoformans* var. *grubii* and 112 proteins of *C. neoformans* var. *neoformans* that share no homology ( $P$  value  $> 10^{-5}$ ) to any proteins in the other variety (Table S2). This apparent discrepancy between the number of variety-specific genes might be due in part to differences in the size of telomeric and sub-telomeric regions available from the two sets, as *C. neoformans* var. *grubii* variant-specific genes map predominantly to chromosomal extremities (Figure S1). Interestingly, we also identified variety-specific transcription factors in *C. neoformans* var. *grubii* (CNAG\_07370 or CNAG\_03745), suggesting the existence of specific regulons. In addition, protein sequence clustering analysis<sup>22</sup> (see methods) revealed the presence of 122 and 129 protein families of at least 2 members in *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*, respectively. In both varieties, the largest variety-specific protein families were related to transposable or retrotransposable elements (Table S3). Finally, reciprocal BLAST (Bdbh) analysis identified 6,341 couples of orthologous proteins between these organisms (Fig. 1A, Table S4). To study the conservation of intron position between the two *C. neoformans* varieties, we restricted our analysis to the 5,712 couples for which the untranslated regions (UTRs) were annotated. As shown in Fig. 1B,  $>99\%$  of the ORF introns are conserved in the exact same position. Nevertheless, we identified 97 ORF introns of variety *neoformans* that were lost in variety *grubii* and conversely 119 introns of variety *grubii* that were lost in variety *neoformans*. The UTR introns appeared to be much less conserved in the 5' UTR and even less in the 3' UTR (Fig. 1B), suggesting that the strong evolutionary pressure that maintains the introns in the same position in the ORF is less marked for these introns. Finally, we compared intron size within the 31,555 couples of orthologous ORF introns and found that 80% differ by  $<2$  nucleotides (Fig. 1C), confirming the high conservation of intron content between the two *C. neoformans* varieties.

**Analysis of alternative splicing.** We first attempted to build a catalogue of all alternative splicing events in *C. neoformans* and then studied their regulation by environmental cues. To evaluate the extent of alternative splicing in *C. neoformans*, we first defined a list of 37,814 constitutive introns (35,822 in ORF, 788 in 3' UTR, and 1204 in 5' UTR) in the coding genes. We then considered the three types of alternative splicing events: IR, A-SS, and ES (see Material and Methods). We identified 1462 alternative splicing events due to A-SS, associated with 1,270 constitutive introns and 7,879 cases of IR. Overall, 59% of the genes have alternatively spliced transcripts in *C. neoformans* var. *neoformans*. Consistent with previous observations in variety *grubii* and other fungi<sup>10,12,13</sup>, IR was the most common form of alternative splicing, followed by A-SS (Fig. 2A). Few cases of ES were observed (Fig. 2A).

Notably, the number of alternative splicing events identified varied between different environmental conditions (Fig. 2B). This finding prompted us to study potential regulation of alternative splicing by environmental cues. We found that more than half ( $n = 4,210$ ) of the IR events and 32% ( $n = 466$ ) of A-SS events were regulated by the growth conditions (as determined by  $\geq 1.5$ -fold change) (Table S5). Semi-quantitative RT-PCR experiments were performed to confirm selected examples of this regulation. Examples of IR and A-SS regulated by the

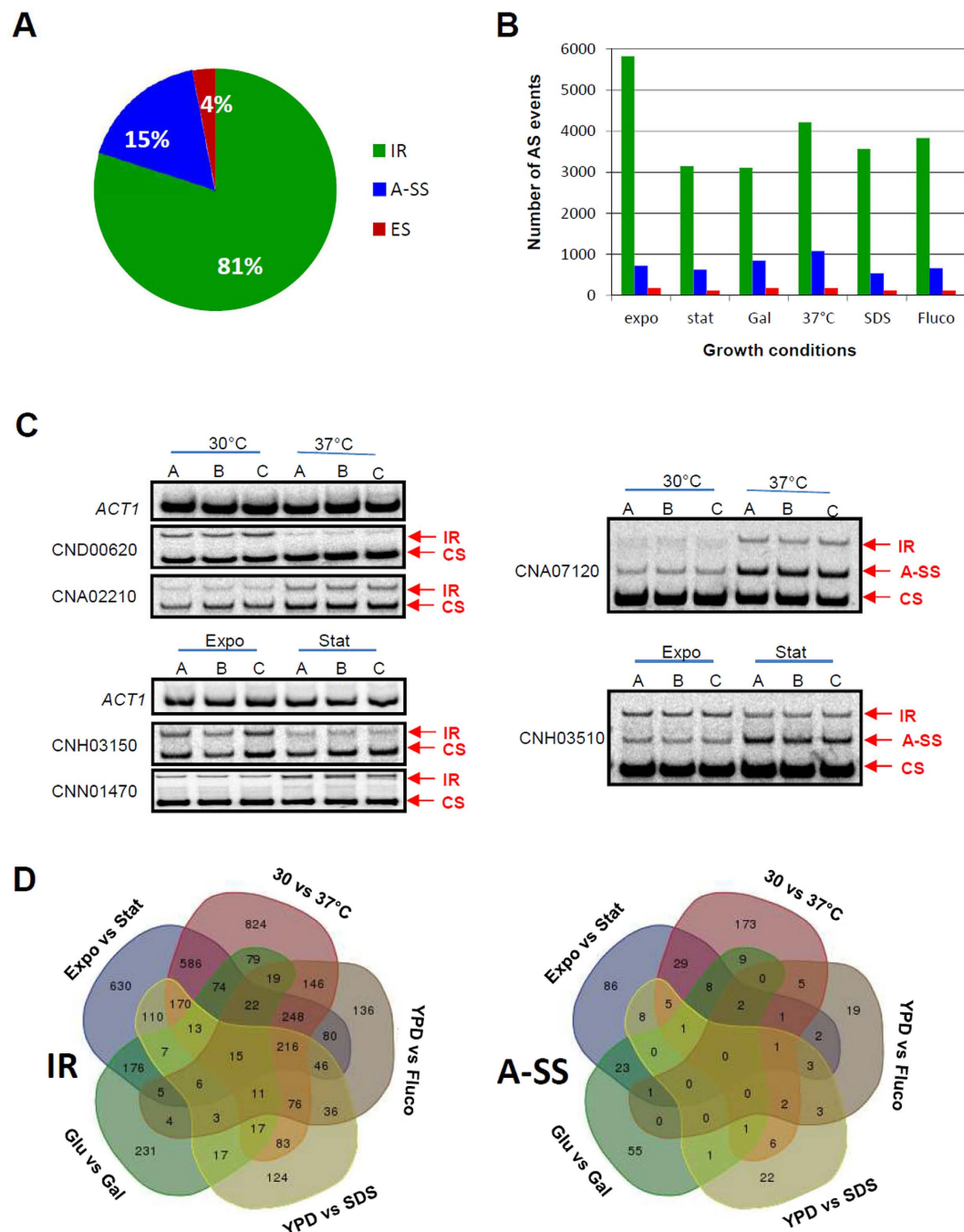


**Figure 1.** Genome comparison between *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*. (A) Position of orthologous genes in *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii* as visualized using SynTView<sup>63</sup>. (B) Right panel. Conservation of intron position between *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii* according to the position of the introns within the genes. The blue bars represent the percentage of var. *grubii* introns conserved in the var. *neoformans* and the red bars represent the percentage of var. *neoformans* introns conserved in the var. *grubii*. Left panel. Size conservation of the ORF introns between the two varieties.

temperature and the growth stage are presented in Fig. 2C. Moreover, the results obtained for the gene CNH03510 illustrated that A-SS and IR at the same locus were not always co-regulated. Interestingly, comparison of the lists of the alternative splicing events regulated in response to a change in one parameter of the growth conditions revealed that many of these regulation changes were specific to the modification (change in temperature, carbon source, or growth stage, or the addition of SDS or fluconazole). Indeed, as shown in Fig. 2D, 46% and 76% of the IR and A-SS regulation, respectively, are comparison specific, suggesting specific alternative splicing regulations in response to environmental modification.

Although IR affects a large number of genes in *C. neoformans*, several lines of evidence suggest that this contributes little to proteomic diversity. First, the mRNAs bearing a retained intron represent only a small fraction of the major type of mRNA of a single gene. As explained in the Material and Methods, an intron was considered to be regulated by IR when at least 5% of the polyadenylated transcripts retained this intron. However, modifying the threshold dramatically altered the number of events. For instance, 4,395 introns were retained in at least 10% of the transcripts in at least one condition, whereas 16,931 introns are retained in at least 1% of the transcripts in at least one condition (Fig. 3A). It is important to note here that these modifications of the threshold only poorly affect the percentages of alternative splicing regulated neither the observed specificity of these regulations (Figure S2). Second, UTR-introns are more likely to be retained than ORF-introns. Thus, the proportion of UTR-introns regulated by IR increases from 7% to 51% when the threshold for the identification of an IR event was modified (Fig. 3C). Third, most IR events are associated with the synthesis of a premature termination codon (PTC)-containing mRNA. Thus, only 31.8% of ORF introns were in-frame with the coding sequence, and 80.1% of these in-frame introns contained a stop codon. Overall, only 7% of the IR events can theoretically result in the production of an alternative protein, underlining the limited potential of IR to regulate proteome diversity in *C. neoformans*. The same is true for alternative splicing events due to 5' and 3' A-SS. The number of A-SS events dramatically decreased when the identification threshold increased, whereas the proportion of events identified in the UTR increased (Fig. 3B,C). Finally, 71% of the A-SS splicing events introduced a frameshift in the coding sequence. Thus, similar to IR, A-SS appears to be a very limited source of proteome diversity in *C. neoformans*, even though some of these frameshifts could result in the production of shorter proteins.

**IR level is mostly independent of nonsense-mediated mRNA decay.** Most alternative splicing events in the ORF likely introduce a PTC into the mRNA sequence. These aberrant mRNAs are expected to be targets of mRNA quality control pathways, thereby avoiding the synthesis of potentially harmful truncated

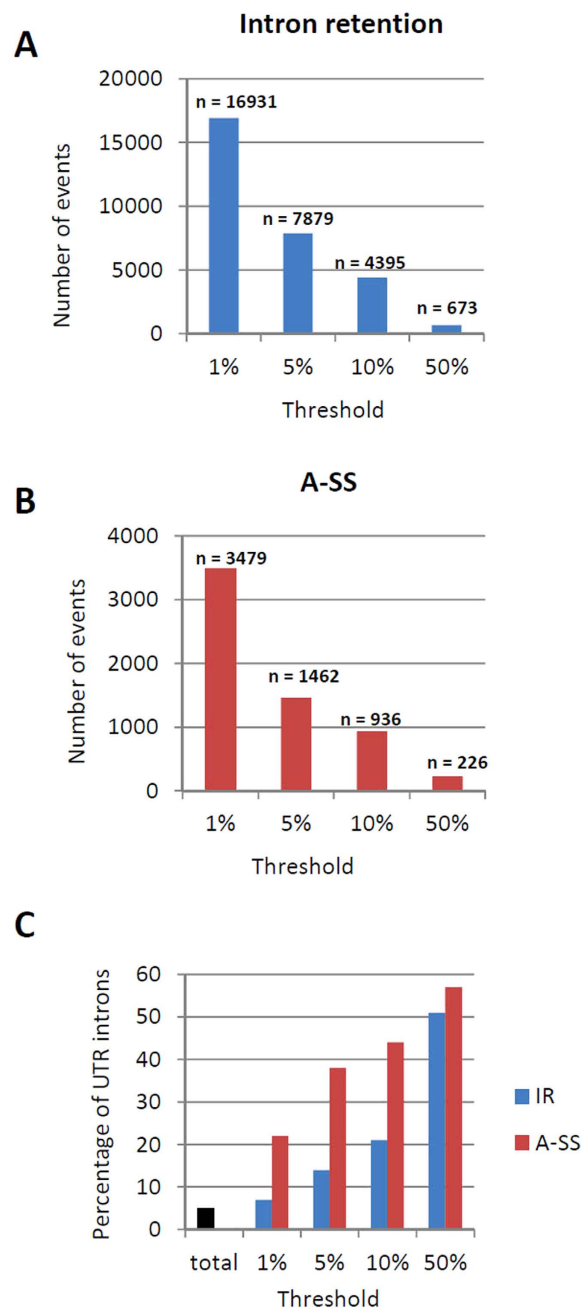


**Figure 2. Regulation of AS by the environment.** (A) Repartition of the different classes of alternative splicing in *C. neoformans*. (B) Number of AS events identified in each growth condition (i.e. exponential phase 30 °C [Expo]; exponential phase 37 °C [37 °C], stationary phase glucose [Stat D]; stationary phase galactose [Stat G]; SDS 0.01% treatment [SDS]; 10 mg /L fluconazole treatment [Fluco]). (C) Examples of RT-PCR analyses of alternative splicing regulation by temperature and stage of growth in *C. neoformans*. Actin mRNA was used as a control. CS, A-SS, and IR stand for constitutive splicing, alternative splice site utilization, and intron retention, respectively. Full-length blots/gels are presented in Supplementary Figure S9. (D) Venn diagrams drawn using lists of AS events in each variety, revealing the specificity of AS regulation by environmental cues.

proteins<sup>23</sup>. Among these, the best studied surveillance pathway is the nonsense-mediated mRNA decay (NMD) pathway, which upon translation recognizes and degrades mRNAs that bear a PTC as a consequence of either a nonsense mutation or a non-productive splicing event<sup>24–27</sup>. Thus, we expected most of the alternative mRNA molecules in *C. neoformans* to be potential NMD targets.

Upf1, Upf2, and Upf3 proteins form the core machinery of the NMD pathway in all eukaryotes studied to date<sup>24</sup>. As expected, we identified a homolog for each *UPF* gene in the genomes of both *C. neoformans* varieties (*UPF1* [loci CNC02960 and CNAG\_01807], *UPF2* [loci CNF01510 and CNAG\_05829], and *UPF3* [loci





**Figure 3. Alternative splicing contributes little to proteome diversity in *C. neoformans*.** (A) Relationship between the number of IR events identified and the threshold used for identification. (B) Relationship between the number of A-SS events identified and the threshold used for identification. (C) Relationship between the number of AS events identified in the UTRs and the threshold used for identification.

CNG03015 and CNAG\_03276]). As shown on the Figure S3A, the *upf* mutant strains did not display obvious growth phenotypes when grown at 30 °C on rich medium. The only clear phenotype observed in all of the *upf*Δ mutants of both varieties was an increased sensitivity to fluconazole. In addition, *upf* mutants in a *C. neoformans* var. *neoformans* background, but not in a *C. neoformans* var. *grubii* background, were temperature-sensitive. Finally, we found that the individual components of the NMD pathway were not required for virulence of *C. neoformans*, as each of the *upf1*Δ, *upf2*Δ, and *upf3*Δ individual or combined mutations resulted in no significant changes in survival patterns in a heterologous host model of cryptococcosis (Figure S3B). In conclusion, as in *S. cerevisiae* and *S. pombe*, depletion of the Upf proteins in *C. neoformans* does not affect cell viability, demonstrating that the NMD pathway is not essential in this yeast.

To get insight into the transcriptomic consequences of *UPF1* deletion, we generated strand-specific RNA-Seq data from RNA samples isolated from a wild-type and the *upf1*Δ mutant strain, both grown to the exponential phase ( $5 \times 10^7$  cell/mL). The paired-end, strand-specific, 100-bp reads were mapped to the *C. neoformans* var.

*neoformans* genome, and statistical analysis identified genes that were significantly up-regulated at least 2-fold ( $n = 411$ ) or down-regulated at least 2-fold ( $n = 338$ ) in the *upf1* $\Delta$  mutant strain, respectively (Table S6). The result is very similar to those previously obtained in mammals, plants, and yeasts, in which about 1% to 10% of the genes were up-regulated upon mutation of the *UPF1* homolog<sup>28–32</sup>. Interestingly, although some of the genes, such as *CPA1* that were previously identified to be up-regulated upon NMD pathway mutation in *S. cerevisiae*<sup>33</sup>, were also up-regulated in *C. neoformans*, the overlap between the two species was quite poor. Thus, 157 of the 411 up-regulated genes upon *UPF1* deletion in *C. neoformans* have an ortholog in *S. cerevisiae*, but only 15 of them are up-regulated in the baker's yeast *nam7* $\Delta$  mutant strain<sup>34</sup>. Moreover, GO-term enrichment analysis of the up-regulated genes did not reveal any pathway particularly altered in the *upf1* $\Delta$  mutant. More specifically, our analysis did not reveal any enrichment of genes involved in amino acid metabolism or in telomere maintenance as previously observed in other organisms<sup>35</sup>. In contrast GO-term analysis of the list of genes down-regulated in the *upf1* $\Delta$  strains revealed a significant enrichment of genes involved in transport. Notably, some genes like *AMF1* (encoding a homolog of a *S. cerevisiae* putative multidrug resistance transporter) and *FLR1* (encoding a homolog of a *S. cerevisiae* fluconazole transporter) were down-regulated 3.6-fold and 4.7-fold, respectively, in the *upf1* $\Delta$  mutant strain<sup>36</sup>. It is tempting to speculate that the down-regulation of these genes could be at least in part responsible for the fluconazole sensitivity of the NMD mutant strains. Additional experiments are needed to confirm this hypothesis.

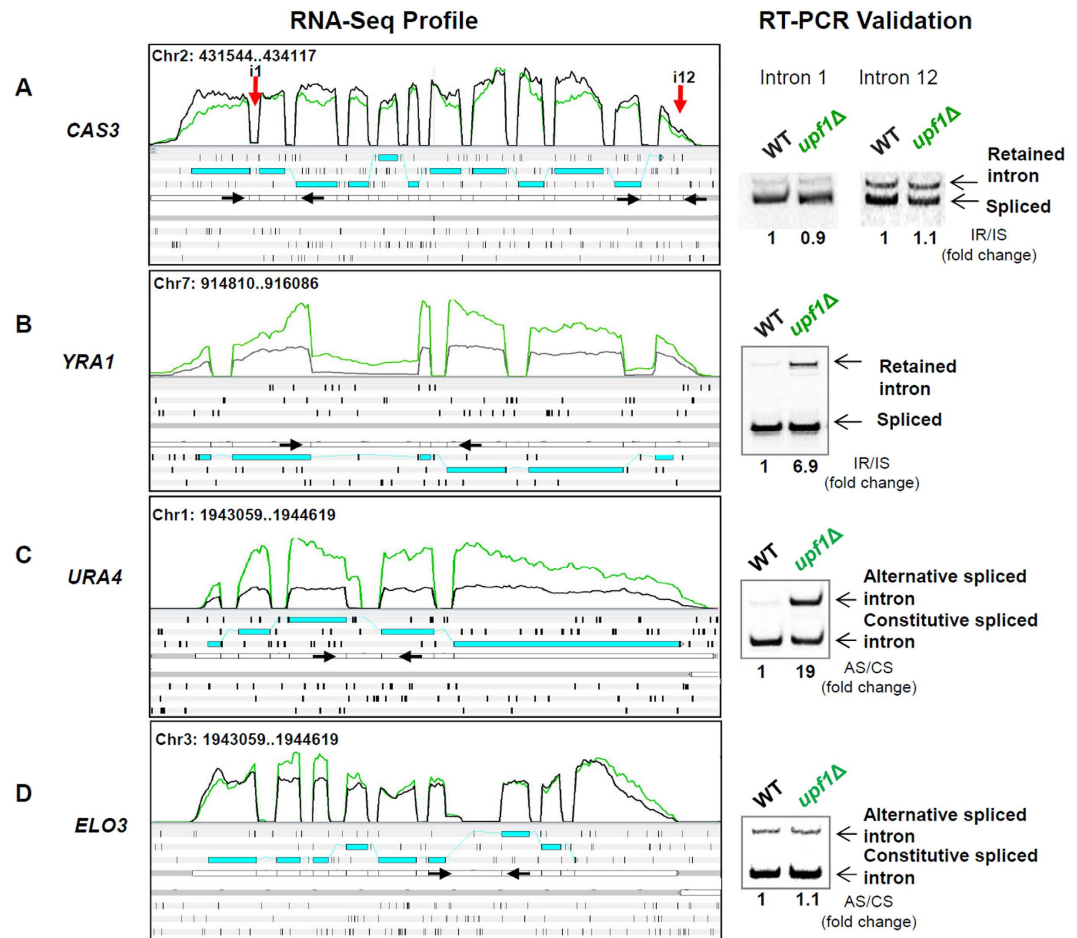
As described above, we expected most intron-retaining mRNAs to be potential NMD targets. One might expect up-regulation of IR upon *UPF1* deletion, as has been shown for other lower eukaryotes, such as *Paramecium tetraurelia*<sup>37</sup> and *S. cerevisiae*<sup>38,39</sup>. We thus studied the influence of NMD pathway deletion on IR level. Here we limited our analysis to the 5,415 introns for which IR could be measured in both wild-type replicates. Strikingly, 98% of the IR events in the ORF and 99% in the UTR were not up-regulated upon *UPF1* deletion, indicating that the vast majority of intron-retaining transcripts are not sensitive to NMD in *C. neoformans*. For instance, we previously reported that the first and last intron of the *CAS3* gene are retained in some of the polyadenylated transcripts, whereas the others are mostly efficiently spliced<sup>20</sup>. The present RNA-Seq-based analysis confirmed that in the wild-type strain, 5.9% and 52.9% of the *CAS3* transcripts retained introns 1 and 12, respectively, whereas none of the other introns were retained at a significant level. However, these IR levels were not affected by the deletion of *UPF1*. Indeed, these levels remained very close to those measured in the wild-type strain (6.1% and 53.9% for introns 1 and 12, respectively). These results were confirmed using semi-quantitative RT-PCR assays as shown in Fig. 4A. We also tested additional loci (CNG00160, CNK02010, and CNA04020) using semi-quantitative RT-PCR assays and obtained similar results, suggesting that NMD has little role in the control of these intron-containing RNA molecules (Figure S4).

Nevertheless, we identified 98 introns in ORFs for which retention was up-regulated in the *upf1* $\Delta$  strains (Table S7). We confirmed the effect of the *UPF1* mutation for some IR events using semi-quantitative RT-PCR. An example of the results obtained for the *YRA1* gene (locus CNG03240) is presented in Fig. 4B. We obtained similar results with the other *upf* mutant strains for both varieties (Figure S5). However, we failed to identify any specific features potentially useful to predict whether the rate of the IR event would be NMD-controlled. Thus, neither the intron size nor the intron position within the gene was predictive of an IR event coupled to NMD-dependent degradation.

In contrast to IR, 52% of the A-SS events identified in the wildtype and which potentially introduce of frameshift in the CDS were shown to be up-regulated in the *upf1* $\Delta$  mutant. More interestingly, 87% ( $n = 297$ ) of A-SS events up-regulated in the *upf1* $\Delta$  mutant strain introduced a frameshift in the coding sequence (Table S7). An example of this category of endogenous NMD target is depicted in Fig. 4C. The *UPF1* deletion revealed the presence of an alternative splicing event of the third intron of *URA4* (CNA07120). RT-PCR experiments using adapted primers confirmed this result. In contrast, most (77%) of the alternative splicing events that did not introduce a frameshift in the coding sequence were not up-regulated in the *upf1* $\Delta$  strain (see representative locus in Fig. 4D). Interestingly, visual examination of the remaining 23% revealed that 31 out of 30 introduced a stop codon to the coding sequence. These data suggest that NMD largely controls the consequences of non-productive A-SS events in *C. neoformans*.

**Link between IR and gene expression regulation.** We analyzed the association between rate of IR and the up- or down-regulation of gene expression. For each growth condition, we arbitrary split the genes in 10 groups, and within the groups the genes were ranked from the lowest to highest level of expression. After eliminating genes that were not expressed in the studied condition ( $\text{fpkm} < 1$ ), we numbered each gene that was regulated by IR within each group. As shown in Fig. 5A, at 30 °C in YPD at exponential phase, the most highly expressed and the least expressed genes tended to be less affected by IR than the genes with moderate expression. A similar pattern was obtained for nearly all conditions (Figure S6). Although the reduced number of IR-regulated genes within the lower expression group is probably due to the criteria used to identify the IR events, the relatively small number of genes affected by IR within the most highly expressed genes suggests a negative correlation between gene expression and intron retention in *C. neoformans*.

Next, we performed a statistical analysis to identify differentially regulated genes in the various growth conditions (Table S8; Figure 5B). For instance, more than a third of the genes ( $n = 2593$ ) were regulated by a change in temperature. Probably due to a general down-regulation of the metabolism together with a slower growth of the cells at 37 °C, ribosome and translation associated-genes were down-regulated at this temperature. Interestingly, some RNA metabolism-associated GO-terms (RNA binding, spliceosomal complex, RNA splicing) were also enriched in the list of the down-regulated genes at this temperature, suggesting that a complex transcriptomic alteration occurs at this temperature. In our analysis, most of the up-regulated genes at 37 °C (71%) have no GO-term annotation, although intracellular and extracellular transport-related genes appeared to be up-regulated. Again these results suggest that the response of *C. neoformans* to the host temperature is not well

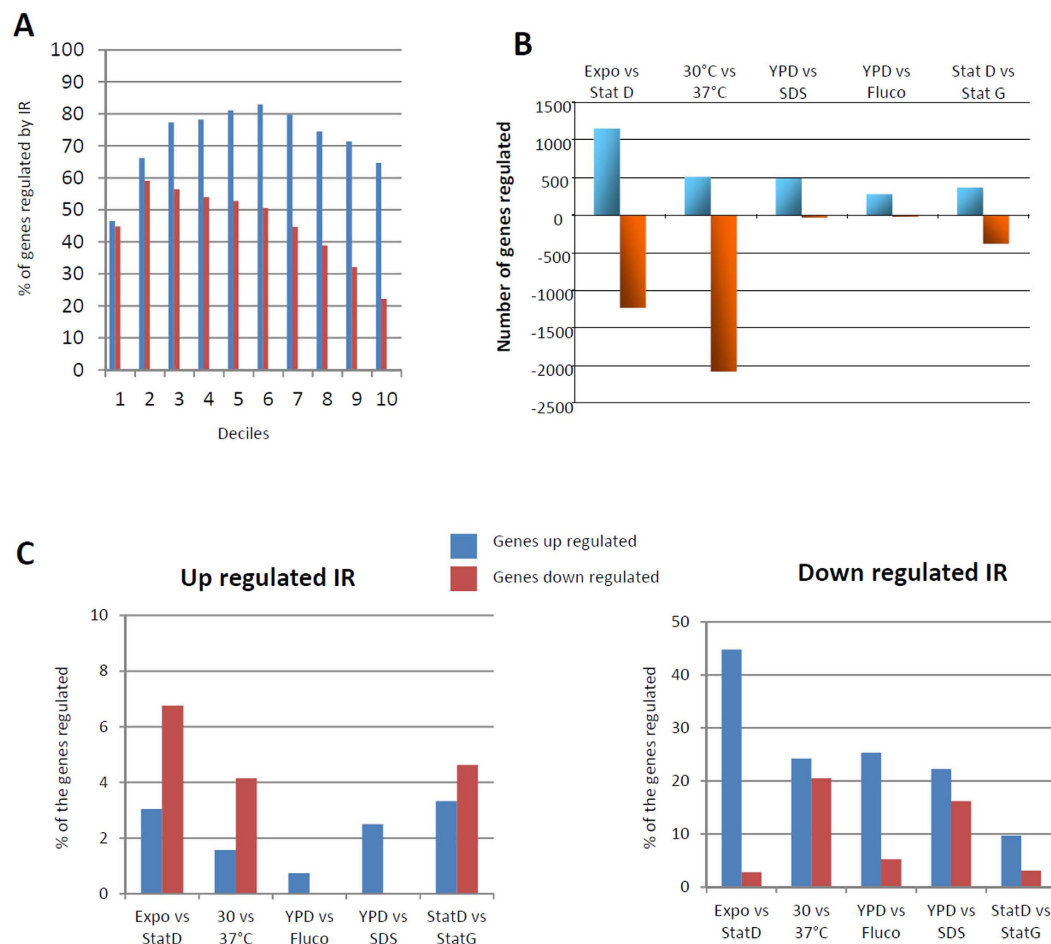


**Figure 4.** RNA-Seq profiles (left) and semi-quantitative RT-PCR validation (right) of representative alternative splicing events that are subject to or immune to NMD. The green and black curves represent the transcription coverage in the *upf1Δ* and wild-type strains, respectively; the blue boxes represent the CDS; and the vertical black bars represent the stop codons found in each frame. Primers (arrows) were chosen on the exons flanking the alternatively spliced introns in order to amplify both the constitutively spliced (CS) and the alternatively spliced (AS) or the unspliced (IR) isoform of each transcript. (A) CAS3 (CNCB01440) is an example of the absence of the IR regulation of upon *UPF1* deletion in *C. neoformans*. The first and the last intron were partly retained (red arrows), and this retention was not altered in the *upf1Δ* strain. (B) YRA1 (CNG03240) is an example of an intron-retaining mRNA up-regulated by *UPF1* deletion. (C) In URA4 (CNA07120) the use of an in-frame alternative 5' splice site introduces a PTC on the mRNA, which is regulated in the *upf1Δ* mutant. (D) In the ELO3 mRNA (CNC00110), the use of an in-frame alternative 5' splice site does not introduce a PTC on the mRNA, and it is equally expressed in the *upf1Δ* mutant as compared to the wild-type strain. Full-length blots/gels are presented in Supplementary Figure S9.

understood. We then studied the relationship between IR level and gene expression regulation. Accordingly, up- and down-regulation of IR in the 6 conditions considered in this study were often associated with the down- and up-regulation of gene expression, respectively (Fig. 5C). For instance, down-regulation of IR was observed for 44% of the up-regulated gene in stationary phase whereas the same observation was made in only 3% of the down-regulated genes. These data clearly suggest the existence of an interdependent relationship between gene expression regulation and IR regulation in *C. neoformans*.

## Discussion

We observed that alternative splicing is common in *C. neoformans* var. *neoformans* and that IR represents its most common manifestation, confirming previous reports in *C. neoformans* var. *grubii* and other fungi<sup>10,21</sup>. However, the level of IR does not appear to be highly influenced by NMD in *C. neoformans*. In this aspect, *C. neoformans* seems to be similar to plants, in which despite a high level of IR only a minority of intron-retaining transcripts is subject to NMD<sup>40,41</sup>. Due to low levels of IR and its small likelihood to generate alternative proteome profiles, one could consider these IR events to be noise due to inefficient intron splicing. However, we previously demonstrated that artificial elimination of the retained introns in a model gene results in enhanced gene expression showing that, at least for this gene, IR is associated with down-regulation of gene expression<sup>20</sup>. Moreover, our present data



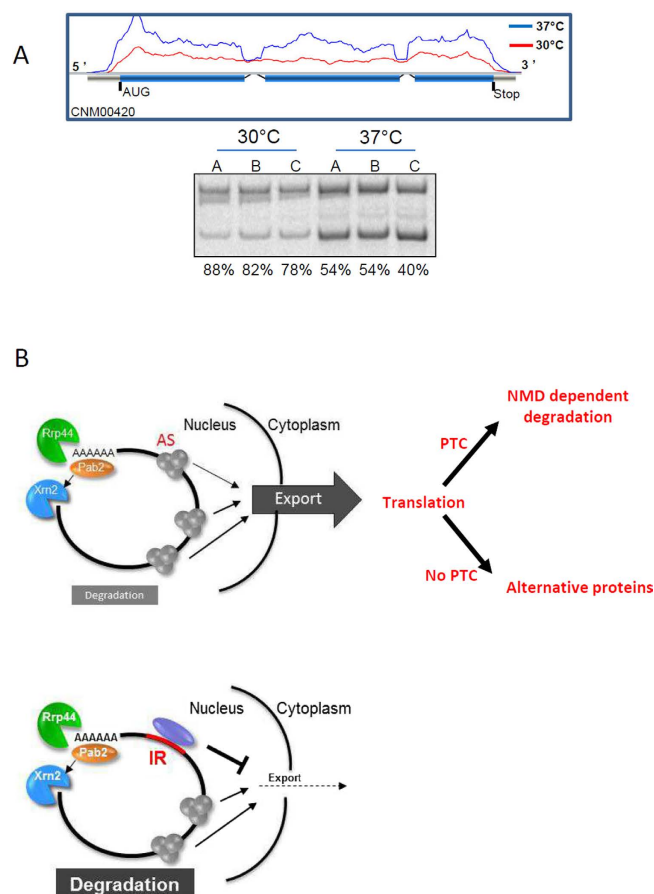
**Figure 5. Intron retention and gene expression regulation is linked.** (A) The most highly expressed genes are less regulated by IR than moderately expressed genes. Genes were separated into deciles according to level of expression at 30 °C in the exponential phase. The percentage of genes regulated by IR in this condition (1% threshold blue; 5% threshold red) is reported. (B) Gene expression regulation in each comparison. (i.e. Exponential phase 30 °C [Expo]; Exponential phase 37 °C [37 °C], Stationary phase glucose [StatD]; Stationary phase galactose [StatG]; SDS 0.01% treatment [SDS]; 10 mg/L fluconazole treatment [Fluco]). (C) Proportion of genes up- or down-regulated when IR is up- or down-regulated. For each comparison, the gene lists for which IR down- or up-regulation was observed were compared with the gene lists for which up- or down-regulation of expression was observed. The percentages of overlap between these lists are reported.

reveal the tight and specific regulation of the IR level by growth conditions and the clear relationship between the regulation of IR level and gene expression, suggesting a controlled and regulated mechanism.

Overall, our data suggest that control of gene expression by IR occurs in *C. neoformans*. Because this regulation is independent of the NMD pathway and thus probably not dependent on translation, one can imagine that this regulation takes place in the nucleus, in which incompletely spliced mRNAs would be retained. One prominent example is the regulation of IR-dependent gene expression by temperature, as shown in Fig. 6A. The CNM00420 gene was transcribed at both 30 °C and 37 °C but was poorly spliced at 30 °C. Thus, the regulation of the expression of this gene at least partly depends on temperature-dependent regulation of IR. The forces inhibiting the export of partially spliced mRNAs are unknown, but they would counteract the forces promoting the export of mRNAs upon splicing<sup>20</sup>. In fact, imaging experiments performed with living plant cells and subcellular fractionation experiments suggested that pre-mRNAs are retained in the nucleus in plant<sup>42,43</sup>. These intron-containing mRNAs could be discriminated from the completely spliced mRNAs by virtue of differences in the set of proteins bound to them<sup>44–48</sup>. In *C. neoformans*, a Spliceosome-Coupled And Nuclear RNAi (SCANR) complex has been shown to mediate the control of transposon expression by targeting transposon transcripts stalled on spliceosomes to degradation in the nucleus<sup>49</sup>. However, the strong siRNA mapping bias to transposons and to sequences sharing similarities with centromeres suggests that this mechanism is restricted to the regulation of transposon expression<sup>49</sup>.

Overall we propose a model (Fig. 6B) in which splicing efficiency is regulated by environmental cues. In this model, IR-regulated mRNAs are exported to the cytoplasm and/or degraded in the nucleus. Thus, IR regulation





**Figure 6. IR-dependent regulation of gene expression in *C. neoformans*.** (A) Example of IR-dependent regulation of gene expression in gene CNM00420. *Top panel.* RNA-Seq profiles as visualized with Artemis. The blue and red curves represent transcription coverage at 30°C and 37°C, respectively; the blue boxes represent the CDS; and the vertical black bars represent the stop codons found in each frame. *Bottom panel.* RT-PCR analysis of intron-retention. Primers (arrows) were chosen on the exons flanking the alternatively spliced introns in order to amplify both the spliced (CS) and unspliced (IR) mRNAs. The experiment was performed on 3 independent replicates and the percentages of retention are given. Full-length blots/gels are presented in Supplementary Figure S9. (B) Model of alternative splicing-dependent regulation of gene expression. *Top panel.* In the case of A-SS, the mRNAs are exported to the cytoplasm. If it contains a PTC, it is degraded by the NMD pathway. In the absence of PTC an alternative protein is produced. *Bottom panel.* If an intron is retained, export from the nucleus is repressed.

represents an additional mechanism in *C. neoformans* to finely tune the level of expression of some genes in order to efficiently adapt to diverse environments.

## Material and Methods

**Strains and culture conditions.** *C. neoformans* strains used in this study originated from the serotype D strain JEC21<sup>50</sup> or serotype A H99<sup>51</sup> and are listed in Table S9. The strains were routinely cultured on YPD medium at 30°C. The bacterial strain *Escherichia coli* XL1-blue (Stratagene) was used for the propagation of all plasmids.

**Construction of deletion and conditional mutants in *C. neoformans* var. *neoformans*.** The *UPF1* (CNC02960), *UPF2* (CNF01510), and *UPF3* (CNBG1750) genes were deleted by biolistic transformation using a disruption cassette constructed by overlapping PCR as previously described<sup>52</sup>. The transformants were then screened for homologous integration as previously described. The plasmid pNAT used to amplify the *NAT* selective marker was kindly provided by Dr. Jennifer Lodge (Saint Louis University School of Medicine). The plasmid pPZP-NEO1 used to amplify the *NEO* selective marker was kindly provided by Dr. Joseph Heitman (Duke University). These cassettes were constructed using a strategy previously applied to *Neurospora crassa* deletion cassettes<sup>53</sup>. All primer sequences used are provided in Table S10. We constructed at least two independent strains for each gene *C. neoformans* var. *neoformans* deletion mutant and analyzed their phenotypes. For each gene, independently constructed mutants exhibited same phenotypic traits.

**Gene disruption of *UPF1*, *UPF2*, and *UPF3* in *C. neoformans* var. *grubii*.** For construction of the serotype A *upf1*Δ, *upf2*Δ, and *upf3*Δ mutants, *UPF1* (CNAG\_01807), *UPF2* (CNAG\_05829), and *UPF3*

(CNAG\_03276) were disrupted in the *C. neoformans* var. *grubii* H99S strain by using the modified split marker/double joint PCR with primers listed in Table S10 as previously reported<sup>54</sup>. The two split gene disruption cassettes were introduced into the H99S strain by biolistic transformation as previously described<sup>55</sup>. For construction of the *upf1Δ upf2Δ* double mutant, the *UPF2* gene was disrupted in the *upf1Δ* mutant with the *NEO* selection marker. For construction of the *upf2Δ upf3Δ* and *upf1Δ upf3Δ* double mutants and the *upf1Δ upf2Δ upf3Δ* triple mutant, *UPF3* was disrupted in the *upf1Δ*, *upf2Δ* or *upf1Δ upf2Δ* mutants with the hygromycin B-resistant marker (*HYG*). The correct genotype of each mutant was verified by Southern blot analysis using a gene-specific probe amplified with the L1/PO primer pair listed in Table S10. We constructed at least two independent strains for each gene *C. neoformans* var. *grubii* deletion mutant and analyzed their phenotypes. For each gene, independently constructed mutants exhibited same phenotypic traits.

**RNA extraction and sequencing.** Total RNA was extracted from *C. neoformans* var. *neoformans* cells grown under various conditions using a previously described protocol<sup>56</sup>. We performed each extraction experiment in independent triplicates for the re-annotation of the genome and the alternative splicing analysis in wild-type strain (JEC21). For the comparison between the wild-type (JEC21) and the *upf1Δ* mutant strain (NE579), we performed each extraction experiment in independent duplicates. For high-throughput sequencing, strand-specific, paired-end cDNA libraries were prepared from 10 μg of total RNA using the Illumina mRNA-Seq-Sample Prep Kit according to manufacturer's instructions. cDNA fragments of ~400 bp were purified from each library and confirmed for quality by Bioanalyzer (Agilent). Then, 100 bp were sequenced from both ends using an Illumina HiSeq2000 instrument according to the manufacturer's instructions (Illumina). Reads were mapped to the genome of strain JEC21 as pairs with Tophat2<sup>57</sup> using the "b2-sensitive" mode with minimum intron length of 5 nucleotides and default settings for other parameters. These alignments were used to correct the *C. neoformans* var. *neoformans* gene structures as previously described<sup>10</sup>.

The number of fragments mapped within the exons of coding genes was counted using the "intersect" function of bedtools2 suite<sup>58</sup>. The average number of fragments mapped within coding gene exons in a single library was 59.1 million (see details in Table S11). Differential expression was investigated using DESeq1 v1.16<sup>59</sup>, DESeq2 v1.4.1<sup>60</sup>, and edgeR v3.6.1<sup>61</sup> with default settings and false discovery rate (FDR) cutoff at 0.05. Only genes with >10 mapped fragments in at least one library were considered. A gene was considered to be significantly differentially expressed when it passed the FDR cutoff in at least 2 of the 3 methods mentioned above. The fold change output from DESeq1 was considered to be the final fold change. RNA-Seq data have been deposited in the NCBI database Bioproject PRJNA272767.

**Orthologous proteins identification.** Reciprocal BLAST<sup>62</sup> (Bdbh) analysis was performed using the sequences of proteins from both varieties. We defined an orthologous couple when both BLASTp *P* values were below  $10^{-5}$ . The positions of orthologous genes in *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii* were visualized using SynTView<sup>63</sup>. The full alignment can be obtained at [http://genopole.pasteur.fr/SynTView/flash/Cryptococcus\\_neoformans\\_grubii\\_H99/SynWeb.html](http://genopole.pasteur.fr/SynTView/flash/Cryptococcus_neoformans_grubii_H99/SynWeb.html).

**Protein family identification.** In order to identify protein families, we performed protein sequence clusterization using CD-HIT<sup>22</sup> (parameters 60%, 60%). An exact duplication of a 62,872-bp fragment at the end of chromosome 12 inverted at the end of chromosome 8 has been previously reported by Fraser and colleagues in 2005<sup>64</sup>. Because this duplicated region is specific to strain JEC21 and is known to have occurred during strain manipulation in the laboratory, the 21 duplicated proteins contained in this DNA region were not considered here as constituting new protein families.

**Comparison of intron positions between *C. neoformans* var. *neoformans* and *C. neoformans* var. *grubii*.** To analyze intron position conservation, the two genome sequences were aligned using Mauve 2.3.1 using default parameters (<http://darlinglab.org/mauve/mauve.html>). We then developed a script to use the position of each extremity of an intron in one variety to infer their theoretical positions in the other variety. We used this tool to classify introns of each strain in 1 of 4 categories: (1) conserved when both predicted positions correspond to intron extremities in the tested variety, (2) cross when the introns in each variety have at least 1 nucleotide overlap, (3) out when the predicted positions are out of the orthologous gene, and (4) lost when no intron is present at the predicted position (Figure S7).

**Analysis of alternative splicing.** Different strategies were used to identify alternative splicing events depending on the type. For IR, we compared the coverage within and just upstream of each constitutive intron as done previously in *C. neoformans* var. *grubii*<sup>10</sup>. For transcripts for which several alternative isoforms were present in the annotation, we chose the most prevalent one when wild-type cells were grown to the exponential phase at 30 °C. Briefly, we measured the coverage in the intron and the coverage within a 2-nt window of the upstream exon (Figure S8). To identify an IR event we used the following 3 criteria: (1) coverage within the upstream exon 2-nt window needed to be at least 10 reads/nt after DESeq1 normalization (DESeq Variance Stabilized Data method); (2) coverage within the intron needed to be at least 3 reads/nt after normalization; and (3) we also limited our analysis to introns displaying at least 5% intron retention (see main text). To consider an IR event as regulated in the *upf1Δ* mutant strain or by a modification of the growth condition, we used 2 additional criteria: (1) the retention level should increase or decrease by at least 1.5-fold with a *P* value < 0.05 as determined by Student *t* test analysis, and (2) the intron should not have been identified as regulated by A-SS. Finally, we eliminated all introns in which an alternative start site has been identified. For alternative splicing due to A-SS, we also limited our analysis to introns for which the coverage within the upstream exon 2-nt window was at least 10 reads/nt after normalization. Moreover an A-SS event needed to be identified by at least 3 reads after normalization in each replicate and represent at least 5% of the constitutive splicing event. An A-SS event was considered to be

regulated when the ratio of reads recognizing the alternative intron compared to the corresponding constitutive one increased or decreased at least 1.5-fold with a  $P$  value  $< 0.05$  as determined by Student  $t$  test analysis.

**RT-PCR analysis.** Total RNA (5  $\mu$ g) was subjected to DNase I treatment (Roche) to eliminate contaminating genomic DNA. A total of 1  $\mu$ g of the DNaseI-treated RNA was then used for reverse-transcription (RT) using the QuantiTect Reverse Transcription (Qiagen) kit. The resultant cDNAs were PCR-amplified in presence of dCTP ( $\alpha^{33}$ P) (Perkin Elmer) with the primers indicated in Table S10. PCR products were resolved on 6% polyacrylamide gel and quantified using a Typhoon 9200 imager and ImageQuant 5.2 software (Molecular dynamics).

**Wax moth killing assay.** To determine the role of NMD in the virulence of *C. neoformans*, independent strains of the *upf1* $\Delta$ , *upf2* $\Delta$ , and *upf3* $\Delta$  single mutants and *upf1* $\Delta$  *upf2* $\Delta$  *upf3* $\Delta$  triple mutants were tested in a *Galleria mellonella* insect model of systemic cryptococcosis. For each group, 15 *G. mellonella* caterpillars in the final instar larval stage, ranging from 200 to 300 mg in body weight, were randomly selected within 7 days from the day of shipment (Vanderhorst, Inc., St Marys Ohio USA). The wild-type (H99S) and mutant strains were grown for 16 hours at 30 °C in YPD medium, washed 3 times with phosphate-buffered saline (PBS), and resuspended in PBS. After cell concentrations were adjusted to 10<sup>6</sup> cells/mL by hemocytometer cell counting, 4  $\mu$ L (4,000 *C. neoformans* cells) were inoculated per larva through the second-to-last prolegs using a 100- $\mu$ L Hamilton syringe with a 10- $\mu$ L needle size. PBS was injected as a non-infection control. Caterpillars were incubated at 37 °C in petri dishes in humidified plastic containers and monitored daily after injection. Caterpillars were considered to be dead when they did not move upon touch or when they displayed a black body color. Caterpillars transforming into pupa during the experiment were censored for statistical analysis. The survival curve was illustrated by Prism 6 (GraphPad) and statistically analyzed by Log-rank (Mantel-Cox) test.

## References

- Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Ann Rev Biochem* **72**, 291–336 (2003).
- Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10**, 741–754 (2009).
- Irimia, M. & Blencowe, B. J. Alternative splicing: decoding an expansive regulatory layer. *Curr Opin Cell Biol* **24**, 323–332 (2012).
- Hui, J. Regulation of mammalian pre-mRNA splicing. *Sci. China. C. Life Sci* **52**, 253–260 (2009).
- Kalsotra, A. & Cooper, T. A. Functional consequences of developmentally regulated alternative splicing. *Nat Rev Genet* **12**, 715–729 (2011).
- Kawashima, T., Douglass, S., Gabunilas, J., Pellegrini, M. & Chanfreau, G. F. Widespread use of non-productive alternative splice sites in *Saccharomyces cerevisiae*. *PLoS Genet* **10**, e1004249 (2014).
- Schreiber, K., Csaba, G., Haslbeck, M. & Zimmer, R. Alternative Splicing in Next Generation Sequencing Data of *Saccharomyces cerevisiae*. *PLoS One* **10**, e0140487, doi: 10.1371/journal.pone.0140487 (2015).
- Neuvéglise, C., Marck, C. & Gaillardin, C. The intronome of budding yeasts. *C R Biol* **334**, 662–670 (2011).
- Wood, V. *et al.* The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871–880 (2002).
- Janbon, G. *et al.* Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. *Plos Genet* **10**, e1004261 (2014).
- Wang, B. *et al.* Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res* **38**, 5075–5087, doi: 10.1093/nar/gkq256 (2010).
- Xie, B.-B. *et al.* Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*. *BMC genomics* **16**, 1–15, doi: 10.1186/s12864-015-1251-8 (2015).
- Grützmann, K. *et al.* Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA Res.* **21**, 27–39 (2014).
- Kim, E., Magen, A. & Ast, G. Different levels of alternative splicing among eukaryotes. *Nucleic Acids Res* **35**, 125–131 (2007).
- Shen, G., Whittington, A., Song, K. & Wang, P. Pleiotropic function of intersectin homologue Cin1 in *Cryptococcus neoformans*. *Mol Microbiol* **76**, 662–676 (2010).
- Rodríguez-Kessler, M. *et al.* Isolation of UmRm75, a gene involved in dimorphism and virulence of *Ustilago maydis*. *Microbiol Res* **167**, 270–282 (2012).
- Wong, S. H. J. & Dumas, B. Ste12 and Ste12-like proteins, fungal transcription factors regulating development and pathogenicity. *Eukaryot Cell* **9**, 480–485 (2010).
- Kabran, P., Rossignol, T., Gaillardin, C., Nicaud, J. M. & Neuvéglise, C. Alternative splicing regulates targeting of malate dehydrogenase in *Yarrowia lipolytica*. *DNA Res.* **19**, 231–244 (2012).
- Dromer, F., Mathoulin-Pélissier, S., Launay, O., Lortholary, O. & the French Cryptococcosis Study, G. Determinants of Disease Presentation and Outcome during Cryptococcosis: The CryptoA/D Study. *PLoS Med* **4**, e21, doi: 10.1371/journal.pmed.0040021 (2007).
- Goebels, C. *et al.* Introns regulate gene expression in *Cryptococcus neoformans* in a Pab2p dependent pathway. *Plos Genet* **9**, e1003686 (2013).
- Loftus, B. *et al.* The genome and transcriptome of *Cryptococcus neoformans*, a basidiomycetous fungal pathogen of humans. *Science* **307**, 1321–1324 (2005).
- Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 658–669 (2006).
- Isken, O. & Maquat, L. E. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* **21**, 1833–1856 (2007).
- Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13**, 703–712 (2012).
- Rehwinkel, J., Raes, J. & Izaurralde, E. Nonsense-mediated mRNA decay: Target genes and functional diversification of effectors. *Trends Biochem Sciences* **31**, 639–646 (2006).
- Neu-Yilik, G. & Kulozik, A. E. NMD: Multitasking between mRNA surveillance and modulation of gene expression. *Adv Genet* **62**, 185–243 (2008).
- Chang, Y. F., Iman, J. S. & Wilkinson, M. F. The nonsense-mediated decay RNA surveillance pathway. *Ann Rev Biochem* **76**, 51–74 (2007).
- Lelivelt, M. J. & Culbertson, M. Yeast Upf proteins required for RNA surveillance affect global expression of the yeast transcriptome. *Mol Cell Biol* **19**, 6710–6719 (1999).
- Rodríguez-Gabriel, M. A., Watt, S., Bähler, J. & P, R. Upf1, an RNA helicase required for nonsense-mediated mRNA decay, modulates the transcriptional response to oxidative stress in fission yeast. *Mol Cell Biol* **26**, 6347–6356 (2006).

30. Yoine, M., Ohto, M. A., Onai, K., Mita, S. & Nakamura, K. The Iba1 mutation of *UPF1* RNA helicase involved in nonsense-mediated mRNA decay causes pleiotropic phenotypic changes and altered sugar signalling in *Arabidopsis*. *Plant J* **47**, 49–62 (2006).
31. Mendell, J. T., Sharifi, N. A., Meyers, J. L., Martinez-Murillo, F. & Dietz, H. C. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36**, 1073–1078 (2004).
32. He, F. *et al.* Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol Cell* **12**, 1439–1452 (2003).
33. Gaba, A., Jacobson, A. & Sachs, M. S. Ribosome occupancy of the yeast CPA1 upstream open reading frame termination codon modulates nonsense-mediated mRNA decay. *Mol Cell* **20**, 449–460 (2005).
34. Malabat, C., Feuerbach, F., Ma, L., Saveanu, C. & Jacquier, A. Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *Elife* **4**, e06722 (2015).
35. Isken, O. & Maquat, L. E. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet* **9**, 699–712 (2008).
36. Gbelska, Y., Krijger, J. J. & Breunig, K. D. Evolution of gene families: the multidrug resistance transporter genes in five related yeast species. *FEMS Yeast Res* **6**, 345–355 (2006).
37. Jaillon, O. *et al.* Translational control of intron splicing in eukaryotes. *Nature* **451**, 359–362 (2008).
38. Sayani, S., Janis, M., Lee, C. Y., Toesca, I. & Chanfreau, G. F. Widespread impact of nonsense-mediated mRNA decay on the yeast intronome. *Mol Cell* **31**, 360–370 (2008).
39. Hossain, M. A., Rodriguez, C. M. & Johnson, T. L. Key features of the two-intron *Saccharomyces cerevisiae* gene *SUS1* contribute to its alternative splicing. *Nucleic Acids Res* **39**, 8612–8627 (2011).
40. Kalya, M. *et al.* Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in *Arabidopsis*. *Nucleic Acids Res* **40**, 2454–2469 (2012).
41. Drechsel, G. *et al.* Nonsense-mediated decay of alternative precursor mRNA splicing variants is a major determinant of the *Arabidopsis* steady state transcriptome. *Plant Cell* **25**, 3726–3742 (2013).
42. Kim, S. H. *et al.* Aberrant mRNA transcripts and the nonsense-mediated decay proteins UPF2 and UPF3 are enriched in the *Arabidopsis* nucleolus. *Plant Cell* **21**, 2045–2057 (2009).
43. Göhring, J., Jacak, J. & Barta, A. Imaging of endogenous messenger RNA splice variants in living cells reveals nuclear retention of transcripts inaccessible to nonsense-mediated decay in *Arabidopsis*. *Plant Cell* **26**, 754–764 (2014).
44. Stutz, F. & Izaurralde, E. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends in Cell Biology* **13**, 319–327 (2003).
45. Bonnet, A., Bretes, H. & Palancade, B. Nuclear pore components affect distinct stages of intron-containing gene expression. *Nucleic Acids Res* **43**, 4249–4261 (2015).
46. Galy, V. *et al.* Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell* **116**, 63–73 (2004).
47. Dziembowski, A. *et al.* Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J* **23**, 4847–4856 (2004).
48. Shiimori, M., Inoue, K. & Sakamoto, H. A specific set of exon junction complex subunits is required for the nuclear retention of unspliced RNAs in *Caenorhabditis elegans*. *Mol Cell Biol* **33**, 444–456 (2013).
49. Dumesic, P. A. *et al.* Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* **152**, 957–968 (2013).
50. Kwon-Chung, K. J., Edman, J. C. & Wickes, B. L. Genetic association of mating types and virulence in *Cryptococcus neoformans*. *Infect Immun* **60**, 602–605 (1992).
51. Perfect, J. R., Ketabchi, N., Cox, G. M., Ingram, C. W. & Beiser, C. L. Karyotyping of *Cryptococcus neoformans* as an epidemiological tool. *J Clin Microbiol* **31**, 3305–3309 (1993).
52. Moyrand, F., Chang, Y. C., Himmelreich, U., Kwon-Chung, K. J. & Janbon, G. Cas3p belongs to a seven member family of capsule structure designer proteins. *Eukaryot Cell* **3**, 1513–1524 (2004).
53. Collopy, P. D. *et al.* High-throughput construction of gene deletion cassettes for generation of *Neurospora crassa* knockout strains. *Methods Mol Biol* **638**, 33–40 (2010).
54. Lee, K. T. *et al.* Distinct and Redundant Roles of Protein Tyrosine Phosphatases Ptp1 and Ptp2 in Governing the Differentiation and Pathogenicity of *Cryptococcus neoformans*. *Eukaryot Cell* **13**, 796–812 (2014).
55. Davidson, R. C. *et al.* Gene disruption by biolistic transformation in serotype D strains of *Cryptococcus neoformans*. *Fungal Genet Biol* **29**, 38–48 (2000).
56. Moyrand, F., Lafontaine, I., Fontaine, T. & Janbon, G. *UGE1* and *UGE2* regulate the UDP-glucose/UDP-galactose equilibrium in *Cryptococcus neoformans*. *Eukaryot Cell* **7**, 2069–2077 (2008).
57. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
58. Quinlan, A. R. & Hall, I. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
59. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
61. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
62. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, doi: 10.1016/s0022-2836(05)80360-2 (1990).
63. Lechat, P., Souche, E. & Moszer, I. SynTVView — an interactive multi-view genome browser for next-generation comparative microorganism genomics. *BMC Bioinformatics* **14**, 1–9, doi: 10.1186/1471-2105-14-277 (2013).
64. Fraser, J. A. *et al.* Chromosomal translocation and segmental duplication in *Cryptococcus neoformans*. *Eukaryot Cell* **4**, 401–406 (2005).

## Acknowledgements

This work was supported by National Research Foundation of Korea grants (2015R1A2A1A15055687) from MEST, the Strategic Initiative for Microbiomes in Agriculture and Food funded by Ministry of Agriculture, Food and Rural Affairs (916006-2) (to Y.S.B). This work was supported by a grant from ANR (2010-BLAN-1620-01 program YeastIntrons) to G.J. We thank Tae-Yup Kim and Anna Floyd for their technical assistance. We thank Cecelia Shertz Wall for editing the manuscript.

## Author Contributions

J.-Y.C., G.J. and Y.-S.B. designed experiments, S.G.-H., F.M., C.P., K.-T.L., E.M. and R.B. performed experiments, D.P., P.L., G.B., C.-C.H., J.H. and G.J. analyzed the data, G.J., S.G.-H. and Y.-S.B. wrote the manuscript, all the authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>



**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Gonzalez-Hilarion, S. *et al.* Intron retention-dependent gene regulation in *Cryptococcus neoformans*. *Sci. Rep.* **6**, 32252; doi: 10.1038/srep32252 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016